DOCUMENT RESUME

ED 180 475 IR 007 949

AUTHOR Markey, Karen
TITLE Online Searching of ERIC: Suggestions for Improvement from the Special Project.
INSTITUTION Syracuse Univ., N.Y. ERIC Clearinghouse on Information Resources.
PUB DATE 14 Oct 79
NOTE 23p.; Paper presented at the ERIC Users' Workshop (Minneapolis, MN, October 14, 1979); For related documents, see IR 007 494-495

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Automatic Indexing; Clearinghouses; *Data Bases; *Information Retrieval; Information Services; *Information Systems; *On Line Systems; Search Strategies

ABSTRACT

Based on a series of studies performed by ERIC/Information Resources staff and consultants which focused on the activities of over 40 experienced online searchers, this report offers the following suggestions for enhancing certain online retrieval system features and making improvements in ERIC vocabulary: (1) the implementation of an online rotated descriptor display, (2) the introduction of an online rotated identifier display, and (3) the use of links between the controlled vocabulary and free text. A brief history of ERIC online services precedes a cursory look at some current ERIC searching practices. In addition, an appendix provides three overheads which compare several elements of ERIC searching with those used by three commercial bibliographic retrieval services presently marketing the ERIC database. (FM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED180475

IR007949

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ERIC/ISE Workshop
October 14, 1979

Online Searching of ERIC: Suggestions for Improvement from the Special Project

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Karen Markey

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Karen Markey
ERIC/IR
Syracuse, New York

ERIC/ISE Workshop
October 14, 1979

Online Searching of ERIC: Suggestions for Improvement from the Special Project

Short History of ERIC Online

The ERIC data base, currently made up of Resources in Education (RIE) and Current Index to Journals in Education (CIJE) was created in 1966 with RIE input only. By 1969 coverage of educational literature published in journals was added by CIJE input. ERIC was conceived as an input/archive system for fugitive literature with two abstracting and indexing publications, RIE and CIJE, and the ERIC Document Reproduction Service (EDRS) for reproduction of microfiche and paper copy of available ERIC documents. The ERIC master tape - the machine-readable records of RIE and CIJE - has been and still is used for printing cumulative indexes.

To use ERIC as a data base, individuals or groups must lease or purchase the tapes from the ERIC Processing and Reference Facility in Bethesda, Maryland, and process them for computer use. The tapes are available at a reasonable price and with only minimum restrictions to encourage wide distribution and use. ERIC, in whole or in part (i.e., either RIE or CIJE), is used by many computer-based bibliographic retrieval systems, academic information centers, and state-wide educational research services. In fact, many of these groups have formatted their own data bases to conform to the ERIC format. The commercial bibliographic retrieval services (BRS, LRS, SDC), however, have reformatted the ERIC data base to fit their system standards since they market other data bases as well and offer special searching features which have made the search service more diversified than ERIC.

Although ERIC sends the same information to each vendor, the reformatting and system differences among the services themselves have resulted both in differences in the way ERIC must be searched and in the results obtained. Since more and more online ERIC searchers are switching back and forth among retrieval systems, e.g. to search other relevant data bases, take advantage of discounts based on volume of use, or use specific retrieval system features, these differences are becoming very important.

ERIC Online Today

Things are changing, however. Central ERIC is now negotiating to receive data base usage reports from BRS, LRS and SDC. Clearinghouses will be conducting training sessions for vendors and users on how to use ERIC online. Completion of the Vocabulary Improvement and Identifier Clean-up Projects will result in an ERIC master tape update. Whenever the ERIC master tape is revised, ERIC tape subscribers and online users are notified by letter and the newsletter ERIC Interchange, respectively. It is then at the discretion of the retrieval services to reload the updated ERIC files.

Overhead 1 (Handout 1) shows master tape data elements codes/CCR worksheet names and their transformation to search qualifiers by the vendors. You can see the differences in the names of the data elements used as search qualifiers by the vendors. Ignoring punctuation, only two fields, AUTHOR (AU) and TITLE (TI), have the same name. Similarly, a listing of the print qualifiers would emphasize differences in the names of data elements across the three commercial systems. Let's expand upon

the boxed portion of this listing which isolates the Basic Index of the three systems.

Overhead 2 shows why different search results are obtained when one searches ERIC on BRS, LRS or SDC. SDC's "neighbor" command produces a display comprised of terms from the basic index and all "search qualifier" fields, e.g. accession number, document type, issue, etc. On the other hand, LRS' comparable command, "expand," calls up a list containing single words from the same fields plus the sponsoring agency, corporate source and note fields. Not only do the data elements contributing to the Basic Index vary, but the command names for Basic Index searching are different.

Overhead 3 shows how differently ERIC's controlled vocabulary is handled on the three systems. SDC's treatment of identifiers precludes retrieving them directly, rather two operations are required. Only BRS allows searchers to retrieve minor descriptors in a single operation. There are thirteen different field names for descriptor types among the three systems from which searchers may choose. The only vendor to offer manipulation of the ERIC Thesaurus, LRS includes basic index postings from single word descriptors in the online Thesaurus display.

Consequently, identical results are rarely obtained when searching the three systems. This is not only due to differences in Basic Indexes but to the recency of ERIC tape revisions and their reloading into the retrieval systems (LRS - 7507; SDC - 7804; BRS - 7909).

Only one of several searches in two of three retrieval systems gives identical results when a single word like Reading is searched. Overhead 4 demonstrates that postings are identical across the two retrieval systems for sponsoring agency. Note, however, that SDC does not include this field in its basic index.

Overhead 3 is only indicative and should not be construed as a definitive statement on the state(s) of the ERIC data base, but it raises the question whether we should talk of one ERIC data base any longer. Online users of ERIC may indeed be using a different version of ERIC when they switch from retrieval system A to retrieval system B, and so on.

Implications from Studying Online Searching of ERIC

To discover something about the use of ERIC online and how it might be improved, a series of studies were performed in late 1978 and early 1979 by ERIC/Information Resources staff and consultants. Over forty experienced ERIC online searchers were interviewed. Conversations with these searchers are summarized in the Special Project Report, Online Searching of ERIC, P. Atherton, principal investigator. Syracuse: ERIC/IR; 1979).

More than half of these searchers volunteered to show our project staff their activity online. Over six hundred search traces were analyzed and findings show that controlled vocabulary terms are used in 68% of ERIC searches. Free text searching is lower if only terms which resulted in offline printed output are considered. Use of retrieval system features related to free text searching, e.g. word adjacency, search saves, online dictionary, truncation, etc., show that searchers employ the fundamental techniques of the system, e.g. selects, combines, prints, which allow them to "get by" with representing the majority of searches expending a minimum of effort and imagination.

Besides a study of these searchers' search strategies, experimental tests were performed to investigate differences between controlled vocabulary and free text searching on LRS. We chose LRS/DIALOG because it was the first to have ERIC online and it has the ERIC Thesaurus online.

Six queries from the ERIC ONTAP Manual (K. Markey and P. Atherton. Syracuse: ERIC/IR; 1978) were used. ERIC Clearinghouse staff from eight clearinghouses cooperated in this study. We found that free text can be a powerful tool for increasing recall and designating specific concepts, but that neither free text nor controlled vocabulary searching alone will retrieve all relevant documents or keep precision figures up consistently.

Based on the above studies and the project staff's own experiences, we can make suggestions for enhancing certain online retrieval system features and making improvements in ERIC vocabulary. For example, we would like to suggest an online rotated descriptor display, something like that now in print but with postings and related term information (see Overhead 5); an online rotated identifier display, probably more important because a printed authority list for ERIC identifiers is not yet available (see Overhead 6); and links between the controlled vocabulary and free text (see Overhead 7). Here, the command, "wrap," would show certain words in context, stripped of non-content words like the, an, a, of, in free text fields such as title or abstract fields of ERIC documents. The command "expand" after a "wrap" command would display descriptors and identifiers assigned to documents which contain the words in context from free text fields. In the example given, three documents contain the words "depression learned helplessness;" expanding from the "wrap" display shows that three descriptors and two identifiers are used in all or combinations of the indexing records of these three documents. Note that such index terms displays might be ranked, that is, the first few index terms might be common to all or almost all documents.

Such displays like the "wrap" and rotated features serve as "wishlists" for the online searching of ERIC. Changes like these could only be brought about if there is interest expressed by ERIC online searchers and ERIC User Groups and if their "demands" are directed to the retrieval systems who could put these features into their systems.

The ERIC system, the retrieval systems, and the user groups are viewed as responsible agents for effecting changes and improvements in searching ERIC online. All we could do is serve as a collecting agency for these suggestions. (See Overheads 8, 9, 10; Handout 2).

ERIC Field Names	ERIC Codes Sheet-Tape	LRS-Search Qualifier	SDC-Search Qualifier	BRS-Search Qualifier
Accession number	0010		/AN	.AN.
Clearinghouse accession number	CH 0011		/CHAN	
Clearinghouse Code		CH=	/CC	.CH.
Author	AUTH 001B	AU=	/AU	.AU.
Geographic source	GEO 004D	CP=	/LO	.GS.
Government level	GOV 004E	GL=	/LO	.GV.
Corporate source	INST 001C	/CS	/OS	.IN.
Sponsoring agency	SPON 0020	/SA	/SPO	.SN.
Title	001A	/TI	/TI	.TI.
Abstract	ABST 002C	/AB	(in BI)	.AB.
Identifier	IDEN 0024	/ID,/ID*, /IF*,/IF	(in BI)	.ID.
Descriptor	DESC 0023	/DE,/DE*, /DF*,/DF	/IT,*/IT, */IW,/IW	.MJ,MN,DE, UJ,UN.
Descriptive note	NOTE 0026	/NT		.NT.
Publication year	PDAT 0017	YR=	FROM_TO_	.YR.
Pagination	PAGE			
Language	LANG 004C	LA=	/LA	.LG.
Journal name	JNL 0032	JO=	/JC	
Legislative authority code		AC=	/LPA	.PA.
Contract/Grant number	CONT 002E GR 002F	CN=	/NU	.NO.
Bureau/Project number	BN 0030	PN=	/NU	.NO.
Report number	REPNO 002D	RN=	/NU	
Corporate Source code	0080	SC=		.IN.
Sponsoring agency code	0084	SC=		.SN.
Document type	PUBTYPE 0013	DT=	/DT	
Group code	0014	GC=		
EDRS price	0025			
Availability level	AVAIL 0031			
Update	0001	UD=	/UP	
Issue	002B	IS=	/IS	

Handout
1. ERIC Codes and Corresponding LRS, SDC, and BRS Search Qualifiers of the ERIC Unit Record. Boxed portion indicates contents of basic index.

ERIC Field Names	ERIC Codes	LRS	SDC	BRS
Accession number	0010		/AN	.AN.
Author	AUTH 001B	AU=	/AU	.AU.
Geographic source	GEO 004D	CP=	/LO	.GS.
Government level	GOV 004E	GL=	/LO	.GV.
Corporate source	INST 001C	/CS	/OS	.IN.
Sponsoring agency	SPON 0020	/SA	/SP	.SN.
Title	001A	/TI	/TI	.TI.
Abstract	ABST 002C	/AB	(IN BI)	.AB.
Identifier	IDEN 0024	/ID,/ID* /IF,/IF*	(IN BI)	.ID.
Descriptor	DESC 0023	/DE,/DE* /DF,/DF*	/IT,*/IT /IW,*/IW	.MJ,MN,DE, UJ,UN.
Descriptive note	NOTE 0026	/NT		.NT.
Language	LANG 004C	LA=	/LA	.LG.

1. ERIC Codes and Corresponding LRS, SDC, and BRS Search Qualifiers of the ERIC Unit Record.

COMMAND AND FIELD NAMES	LRS	SDC	BRS
COMMAND TO SEARCH BASIC INDEX:	SELECT	FIND	..SEARCH
BASIC INDEX CONTAINS SINGLE WORDS FROM:			
FIELD NAME			
TITLE	TI	TI	TI
SPONSORING AGENCY	SA	-	SN
CORPORATE SOURCE	CS	-	IN
ABSTRACT	AB	(AB)	AB
NOTE	NT	-	NT
DESCRIPTOR	DE,DF, DE*,DF*	IT,IW, IT*,IW*	DE,MJ, MN,UJ,UN
IDENTIFIER	ID,IF, ID*,IF*	(ST)	ID
GOVERNMENT LEVEL	-	-	GV
GEOGRAPHIC SOURCE	-	-	GS
COMMAND TO BROWSE BASIC INDEX:	EXPAND	NEIGHBOR	ROOT
COMMAND TO BROWSE ERIC THESAURUS:	EXPAND		

2. DIRECTLY SEARCHABLE BASIC INDEX FIELDS AND RELATED SYSTEM COMMANDS OF ERIC ONLINE.

DESCRIPTOR TYPE	LRS	SDC	BRS
MAJOR BOUND DESCRIPTORS	/DF*	*/IT	.MJ.
MINOR BOUND DESCRIPTORS	-	-	.MN.
MAJOR OR MINOR BOUND DESCRIPTORS	/DF	/IT	.MN,MJ.
MAJOR SINGLE DESCRIPTOR WORDS	/DE*	*/IW	.UJ,MJ.
MINOR SINGLE DESCRIPTOR WORDS	-	-	.UN,MN.
MAJOR OR MINOR SINGLE DESCRIPTOR WORDS	/DE	/IT	.DE.

3A. CONVENTIONS FOR DESIGNATING DIRECTLY SEARCHABLE CONTROLLED VOCABULARY FIELDS IN A SINGLE OPERATION: DESCRIPTORS.

IDENTIFIER TYPE	LRS	SDC	BRS
MAJOR BOUND IDENTIFIER	/IF*	-	-
MINOR BOUND IDENTIFIER	-	-	-
MAJOR OR MINOR BOUND IDENTIFIERS	/IF	-	.ID. (USE ADJ)
MAJOR SINGLE IDENTIFIER WORDS	/ID*	-	-
MINOR SINGLE IDENTIFIER WORDS	-	-	-
MAJOR OR MINOR SINGLE IDENTIFIER WORDS	/ID	-	.ID.

3B. CONVENTIONS FOR DESIGNATING DIRECTLY SEARCHABLE CONTROLLED VOCABULARY FIELDS IN A SINGLE OPERATION: IDENTIFIERS.

BASIC INDEX FIELDS		LRS	SDC	BRS
	CIJE:	08/79	02/79	08/79
	RIE:	08/79	06/79	08/79
TOTAL		29,255	28,156	29,323
TITLE		9,926/TI	9,642/TI	9,958.TI.
ABSTRACT		20,430/AB	?	?
SPONSORING AGENCY		7/SA	-	7.SN.
NOTE		2,008/NT	-	2,004.NT.
CORPORATE SOURCE		792/CS	-	790.IN.
GOVERNMENT LEVEL		-	-	0.GV.
GEOGRAPHIC SOURCE		-	-	0.GS.

4A. FIELD BY FIELD TALLY OF POSTINGS FOR SEARCHING A SINGLE WORD, READING, IN THE BASIC INDEX FIELDS OF ERIC ONLINE (10/9/79).

BASIC INDEX FIELDS		LRS	SDC	BRS
	CIJE:	08/79	02/79	08/79
	RIE:	08/79	06/79	08/79
MAJOR BOUND DESCRIPTOR		1,032/DF*	1,010*/IT	1,034.MJ.
MINOR BOUND DESCRIPTOR		-	-	1,878.MN.
MAJOR OR MINOR BOUND DESCRIPTOR		2,906/DF	2,867/IT	2,911.MJ,MN.
MAJOR SINGLE DESCRIPTOR WORD		16,379/DE*	15,895*/IW	16,399.MJ,UJ.
MINOR SINGLE DESCRIPTOR WORD		-	-	14,397.MN,UN.
MAJOR OR MINOR SINGLE DESCRIPTOR WORD		21,957/DE	21,466/IW	22,010.DE.
MAJOR BOUND IDENTIFIER		0/IF*	-	-
MAJOR OR MINOR BOUND IDENTIFIER		0/IF	-	-
MAJOR SINGLE IDENTIFIER WORD		670/ID*	-	-
MAJOR OR MINOR SINGLE IDENTIFIER WORD		1,151/ID	-	1,156.ID.

4B. FIELD BY FIELD TALLY OF POSTINGS FOR SEARCHING A SINGLE WORD, READING, IN THE BASIC INDEX FIELDS OF ERIC ONLINE (10/9/79).

? ROTATE MUSIC/DE

REF	INDEX-TERM	ITEMS	RT
T1	MUSCULAR STRENGTH	150	6
T2	MUSEUMS	273	9
T3	-MUSIC-	946	21
T4	MUSIC ACTIVITIES	401	12
T5	APPLIED MUSIC	49	5
T6	ASIAN MUSIC		1
T7	MUSIC APPRECIATION	222	2
T8	CHORAL MUSIC	51	8
T9	MUSIC EDUCATION	1197	11
T10	MUSIC FACILITIES	38	7
T11	ORIENTAL MUSIC	15	1
T12	MUSIC READING	80	1
T13	MUSIC TEACHERS	177	1

-MORE-

5. EXAMPLE OF AN ONLINE ROTATED DESCRIPTOR DISPLAY.

? ROTATE H/ID

REF	INDEX-TERM	ITEMS	RT
I1	FOUR H	1	
I2	FOUR H CAREER EXPLORATION PROGRAM	1	
I3	FOUR H CENTER	1	
I4	FOUR H CLUB	1	
I5	FOUR H CLUB AGENTS	1	
I6	FOUR H CLUBS	65	
I7	FOUR H PROGRAM	3	
I8	FOUR H PROGRAMS	2	
I9	4 H	1	
I10	4 H PROGRAMS	4	
I11	4 H TEEN INTERN REPORT	1	
I12	H R 11258	1	
I13	H R 12257	2	
I14	H R 13266	1	

-MORE-

6. EXAMPLE OF AN ONLINE ROTATED IDENTIFIER DISPLAY.

```
? WRAP DEPRESSION/TI

  REF                          INDEX-TERM                          ITEMS  TT
  W1                           DEPRESSION LEARNED HELPLESSNESS       3    5
  W2                           DEPRESSION OPPRESSED                  1
  W3                           DEPRESSION RECOVERY HIGHER            1
  ...                          ...
  W10                          DEPRESSION SPEECH INITIAL             1
  W11    IRRATIONAL BELIEFS    DEPRESSION                            1
  W12  LEARNED HELPLESSNESS    DEPRESSION                            2    5
  W13  LEARNED HELPLESSNESS    DEPRESSION CLINICAL MODEL             1    5
                                      -MORE-

? EXPAND W1

  REF                          INDEX-TERM                          ITEMS  RT
  T1                           PSYCHOPATHOLOGY                     322    9
  T2                           PSYCHOLOGICAL STUDIES               2024   5
  T3                           PSYCHOLOGICAL PATTERNS              1074   21
  T4                           LEARNED HELPLESSNESS                44
  T5                           DEPRESSION                          201
```

7. Imagined Transaction of Linkage between Free Text Title Terms and Controlled Vocabulary.

1. ESTABLISH LINKS BETWEEN USER GROUPS, ANSI, NFAIS, ASIDIC, FOR DEVELOPING MULTIPLE DATA BASE STANDARDS.

2. DEVELOP STANDARDS FOR NAMING OF DATA ELEMENTS, CONTENTS OF ERIC BASIC INDEX.

3. REVIEW DIFFERENCES IN ACCESS TO DATA ELEMENTS IN ERIC ONLINE.

4. STUDY SEARCH OBJECTIVES TO SEE HOW THEY RELATE TO REQUIREMENTS FOR SYSTEM FEATURES AND DATA ELEMENTS.

5. ASSESS METHODS FOR IMPROVING SEARCHERS' COMPETENCIES.

6. SUGGEST NEW SYSTEM FEATURES.

7. DEVELOP A REVIEW BOARD FOR ENDORSING PRINTED AND ONLINE USER AIDS.

8. CALL FOR COORDINATION AMONG RETRIEVAL SYSTEMS AS TO RELOADING OF ERIC MASTER TAPE UPDATES.

1. PROVIDE LINKS BETWEEN BASIC INDEX AND THE CONTROLLED VOCABULARY.

2. COORDINATE AND ANNOUNCE PLANNED RELOADING OF ERIC FILES.

3. CONFORM TO ERIC STANDARD OF ORDER OF PRINTING DATA ELEMENTS.

4. COOPERATE WITH USER GROUPS, ANSI, NFAIS, ASIDIC, IN DEVELOPMENT AND IMPLEMENTATION OF STANDARDS.

5. ENHANCE ONLINE ACCESS TO ERIC THESAURUS.

6. IMPLEMENT NEW FEATURES TO AID ERIC ONLINE SEARCHERS UPON RECOMMENDATION AND REVIEW BY USER GROUPS.

1. SEEK ECONOMIC APPROACH TO DISTRIBUTION OF ERIC MASTER TAPE UPDATES.

2. DEVELOP HIGH QUALITY USER AIDS FOR ONLINE SEARCHING.

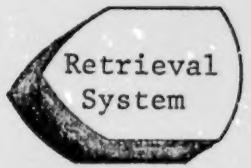
3. FIND ADDITIONAL WAYS OF LIMITING EXCESSIVE OUTPUT.

4. STUDY FREE TEXT AND CONTROLLED VOCABULARY SEARCHING.

5. DEVELOP PRECISE RULES FOR INDEXING AND ABSTRACTING EXPRESSIONS OF ACTION.

6. TEST TO CHECK SAVINGS IN TIME AND COST USING VARIOUS SEARCH SAVES. MAKE AVAILABLE PUBLIC SEARCH SAVES.

1. Establish links between User Groups, ANSI, NFAIS, ASIDIC, for developing standards.
2. Develop standards for naming of data elements, contents of ERIC basic index.
3. Review differences in access to data elements in ERIC online.
4. Study search objectives to see how they relate to requirements for system features and data elements.
5. Assess methods for improving searchers' competencies.
6. Suggest new system features.
7. Develop a review board for endorsing printed and online user aids.
8. Call for coordination among retrieval systems as to reloading of ERIC master tape updates.



1. Provide links between basic index and controlled vocabulary.
2. Coordinate and announce planned reloading of ERIC files.
3. Conform to ERIC standard for order of printing data elements.
4. Cooperate with User Groups, ANSI, NFAIS, ASIDIC, in development and implementation of standards.
5. Enhance online access to ERIC Thesaurus.
6. Implement new features to aid ERIC online searchers upon recommendation and review by User Groups.

1. Seek economic approach to distribution of ERIC master tape updates.
2. Develop high quality user aids for online searching.
3. Find additional ways of limiting excessive output.
4. Study free text and controlled vocabulary searching.
5. Develop precise rules for indexing and abstracting of expressions of action.
6. Test to check savings in time and cost using various search saves. Make available Public Search Saves.

Handout
2. List of suggestions for improvements in searching ERIC online: ERIC User Group, Retrieval System, ERIC System.